

ME 705 A - Inferência Bayesiana
Primeiro semestre de 2023
Lista de Exercícios V

OBS: A menos que o contrário seja mencionado, nos exercícios você deve considerar uma amostra aleatória $X_1|\theta, \dots, X_n|\theta$ de $X|\theta$.

OBS: Em todas as questões envolvendo análise de dados, reflita, criticamente, se o(s) modelo(s) sugerido(s) são, de fato, os mais apropriados. Sinta-se livre para considerar modelos e/ou prioris que não aqueles mencionados nas questões. Também, verifique a adequabilidade do ajuste do(s) modelo(s) considerado(s), usando pelo menos uma das metodologias vistas, bem como comparando os modelos, quando mais de um for ajustado.

1. Resolva todos os exercícios deixados em sala.
2. Considere os exercícios 4, 10 e 11, da Lista II; 4 da Lista III, 2 da Lista IV. Sob a família conjugada de prioris e/ou sob a priori de Jeffreys (sempre que for possível obter cada uma delas), encontre as distribuições condicionais completas de cada parâmetro (univariadamente) condicionado nos demais. Caso não seja possível de obter nem a família conjugada nem a priori de Jeffreys, escolha uma priori e encontre as referidas distribuições.
3. Para as distribuições mencionadas no exercício 2 desta Lista, implemente, no INLA e/ou stan (rstan, ou stan_glm ou brms) do R, a obtenção das distribuições à posteriori.
4. Para os exemplos apresentados nos dois conjuntos de slides sobre algoritmos MCMC, encontre as distribuições condicionais completas sob as famílias de prioris adotadas. Separe por tipo de parâmetro, ou seja, obtenha as condicionais completas para todo o vetor β e, para os outros parâmetros, se for o caso, univariadamente.
5. Seja $X|r \sim \text{Weibull}(r, \lambda)$, λ conhecido. Sua fdp é dada por

$$p(x|r) = \frac{r}{\lambda} \left(\frac{x}{\lambda}\right)^{r-1} e^{-\left(\frac{x}{\lambda}\right)^r} \mathbb{1}_{(0, \infty)}(x)$$

Responda os itens:

- a) Obtenha a priori de Jeffreys para r e a respectiva posteriori. A priori e/ou a posteriori são próprias?
- b) Assum a priori que $p(r) \propto r^{\alpha-1} e^{-r/\delta} \mathbb{1}_{(0, \infty)}(r)$. Encontre a respectiva posteriori.
- c) Implemente no R, usando a função “integrate”, um algoritmo para obter uma aproximação numérica para as posterioris dos itens a), b).
- d) Implemente no R, usando a função “integrate”, algoritmos para obter aproximações numéricas para o EAP, e o EPAP, para cada uma das posterioris dos itens a), b).
- e) Implemente no R, usando a função “optim”, algoritmos para obter aproximações numéricas para o MAP e o MeAP, para cada uma das posterioris dos itens a), b).
- f) Implemente no R, usando a função “integrate”, um algoritmo para obter uma aproximação numérica para o fator de Bayes para testar as hipóteses $H_0 : r \leq r_0$ vs $H_1 : r > r_0$, r_0 conhecido, para cada uma das posterioris dos itens a), b).

g) Considere as hipóteses $H_0 : r = r_0$ vs $H_1 : r \neq r_0$, r_0 conhecido. Assuma à priori que:

$$p(r) = [\gamma \mathbb{1}_{\{r_0\}}(r) + (1 - \gamma)p_1(r)\mathbb{1}_{\Theta_1}(r)], \Theta_1 = (0, \infty) - r_0,$$

quem que $p_1(r)$ é a priori dada no item b). Implemente no R, usando a função “integrate”, um algoritmo para obter uma aproximação numérica para o fator de Bayes para testar as hipóteses em questão.

6. Para a questão anterior, considere λ desconhecido e os dados do exemplo da potência das turbinas de avião (considerando e desconsiderando os grupos). Implemente no INLA e/ou stan (rstan, ou stan_glm ou brms) a obtenção das posterioris, com prioris de sua escolha.
7. Considere os dados sobre a idade de ocorrência de menarca de garotas de Varsóvia (arquivo meninas.txt). Em cada coluna tem-se, respectivamente, o número de garotas que apresentaram menstruação, o número de garotas entrevistadas e a idade média do grupo. Primeiramente, realize uma análise descritiva adequada. Posteriormente, assumindo prioris do tipo $\beta_i \sim N(0, \psi)$ (independentes), em que ψ deve ser escolhido por você, implemente no INLA e/ou stan (rstan, ou stan_glm ou brms) o ajuste bayesiano do modelo de regressão logístico com distribuição binomial. Ou seja, $Y_i | (\beta_0, \beta_1) \sim \text{Binomial}(m_i, p_i)$, m_i : é o número de garotas entrevistadas com idade i , Y_i : é o número de meninas que já apresentaram menarca com idade i e $\text{logit}(p_i) = \beta_0 + \beta_1(x_i - \bar{x})$, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $n = 25$ e x_i : é a idade média do grupo i . Comente sobre a convergência do algoritmo, o comportamento das distribuições à posteriori, e sobre a significância dos parâmetros do modelo. Não se esqueça de fornecer estimativas pontuais, os desvios-padrão à posteriori e intervalos de credibilidade (incluindo o intervalo HPD). O aumento na probabilidade de menarca aumenta, substancialmente, com o aumento em uma ano na idade média? Além disso, faça um gráfico com as proporções observadas, esperadas e os respectivos intervalos HPD, sob o modelo, em função da idade (à semelhança do exemplo dos besouros, visto em sala). Verifique a qualidade de ajuste do modelo e, se for o caso ajuste outros modelos (ou seja, usando outras funções de ligação mas, mesma distribuição), comparando os respectivos ajustes. Forneça suas conclusões com o modelo que melhor se ajustou aos dados. Para mais detalhes veja as páginas 123 e 124 do livro do Prof. Gilberto Paula, Modelos de Regressão com apoio computacional.
8. Os dados abaixo são relativos ao número de bactérias sobreviventes em amostras de um produto alimentício segundo o tempo (em minutos) de exposição do produto à uma temperatura de $300^\circ F$. Para maiores detalhes veja página 61 do livro do Prof. Gilberto Paula, modelos de Regressão com apoio computacional. Primeiramente, realize uma análise descritiva adequada. Posteriormente, assumindo prioris do tipo $\beta_i \sim N(0, \psi)$ (independentes), em que ψ deve ser escolhido por você, implemente no INLA e/ou stan (rstan, ou stan_glm ou brms) o ajuste bayesiano do modelo de regressão Poisson com ligação log. Ou seja, $Y_i | (\beta_0, \beta_1) \sim \text{Poisson}(\mu_i)$ e $\ln(\mu_i) = \beta_0 + \beta_1(x_i - \bar{x})$, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$, $n = 12$ e x_i : é o tempo de exposição (em minutos). Comente sobre a convergência do algoritmo, o comportamento das distribuições à posteriori, e sobre a significância dos parâmetros do modelo. Não se esqueça de fornecer estimativas pontuais, os desvios-padrão à posteriori e intervalos de credibilidade (incluindo o intervalo HPD). O aumento em uma unidade do tempo de exposição implica numa diminuição significativa do número de bactérias? Além disso, faça um gráfico com o número de bactérias sobreviventes observado, esperado e os respectivos intervalos HPD, sob o modelo, em função do tempo de exposição (à semelhança do exemplo dos besouros, visto em sala). Verifique a qualidade de ajuste do modelo e, se for o caso ajuste outros modelos (ou seja, usando outras distribuições para dados de contagem), comparando os respectivos ajustes. Forneça suas conclusões com o modelo que melhor se ajustou aos dados.

Número	175	108	95	82	71	50	49	31	28	17	16	11
Tempo	1	2	3	4	5	6	7	8	9	10	11	12

9. Considere os resultados de um estudo sociológico desenvolvido na Austrália com 146 estudantes de 8^a série e ensino médio com o objetivo de comparar a ausência na escola segundo os seguintes fatores (covariáveis): ano que o estudante está cursando (1: 8^a série, 2: 1^o ano do ensino médio, 3: 2^o ano do ensino médio, 4: 3^o ano do ensino médio), etnia (0: aborígine, 1: não aborígine), desempenho escolar (0: insuficiente, 1: suficiente) e sexo (0: masculino, 1: feminino) e o número de faltas (variável resposta). Os dados estão disponíveis no arquivo quine.dat. Para as análises, considere somente os efeitos principais (ou seja, NÃO considere interações de nenhuma ordem). A pergunta é como as covariáveis afetam a resposta. assumindo prioris do tipo $\beta_i \sim N(0, \psi)$ (independentes), em que ψ deve ser escolhido por você, implemente no INLA e/ou stan (rstan, ou stan.glm ou brms) o ajuste bayesiano o ajuste bayesiano de (algum) modelo(s) apropriado(s). Verifique a qualidade de ajuste do modelo inicial e, se for o caso ajuste outros modelos, e compare os ajuste. Forneça suas conclusões com o modelo que melhor se ajustou aos dados. Glossário: 1. etnia (A:aborígine, N:não aborígine), 2. sexo (M:masculino, F: feminino), 3. ano (F0:8^a série, F1:1^o ano ensino médio, F2:2^o ano ensino médio, F3:3^o ano ensino médio), 4. desempenho (SL:baixo, AL:normal)